DOCUMENT RESUME

ED 261 698 IR 051 269

AUTHOR Bollinger, William A.; And Others
TITLE Post-Processing of Bibliographic Citations from DOE/RECON, NASA/RECON, and DOD/DROLS.
INSTITUTION California Univ., Livermore. Lawrence Livermore Lab.
SPONS AGENCY Department of Energy, Washington, D.C.
REPORT NO UCRL-89995-Rev-1
PUB DATE Aug 84
CONTRACT W-7405-ENG-48
NOTE 18p.; Paper presented at the International Online Information Meeting (8th, London, England, December 4-6, 1984).
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Bibliographic Coupling; Citation Indexes; *Citations (References); *Computer Software; Correlation; *Databases; *Data Processing; *Information Networks; *Information Retrieval; Information Systems; *Online Systems
IDENTIFIERS *Bibliographic Services; Gateway Systems

ABSTRACT

An interactive, self-guided program for the joint post-processing of bibliographic citations from the federal information centers of the Department of Energy (DOE), the Department of Defense (DOD), and the National Aeronautics and Space Administration (NASA) is described. The program, presently installed on the Intelligent Gateway Processor of the Technology Information System (TIS/IGP) at the Lawrence Livermore National Laboratory, is under evaluation by the TIS user community from remote terminals by telephone dial-up over TYMNET, and the ARPA computer network. Users are individually authorized for access to specific information centers, and use standard commands for the downloading, compilation, and online review of citations in a common format. Previously reported post-processing capabilities have been expanded, permitting: (1) online citation review, categorization, and addition of new data elements; (2) disassembly and re-assembly of citations; (3) statistical analysis of data field contents; (4) cross-correlation of data field contents; and (5) concordance generation. Additionally, the new two-pass interpreter for the post-processing program permits: the transformation of abbreviated data field names into English names preferred by each agency, the statistical analysis of the density and completeness of data fields in selected sets of bibliographic citations, the elimination of redundant citations, and trend analysis. References are included. (Author/THC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

UCRL-89995 Rev. 1
PREPRINT

ED261698

POST-PROCESSING OF BIBLIOGRAPHIC CITATIONS FROM DOE/RECON, NASA/RECON, and DOD/DROLS

William A. Bollinger
Viktor E. Hampel
Isom Harrison
Thomas P. Murphy

Prepared for presentation to
Eighth International Online Information Meeting
London, England
December 4-6, 1984

August 1984

BEST COPY AVAILABLE

IR051269

DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

BEST COPY AVAILABLE

UCRL-89995
Rev. 1

POST-PROCESSING OF BIBLIOGRAPHIC CITATIONS FROM DOE/RECON, NASA/RECON, and DOD/DROLS*

William A. Bollinger
Viktor E. Hampel
Isom Harrison
Thomas P. Murphy

Technology Information System
Lawrence Livermore National Laboratory
P. O. Box 808, L-275
Livermore, CA 94550

ABSTRACT

We have developed an interactive, self-guided program for the joint post-processing of bibliographic citations from the federal information centers of the Department of Energy (DOE), the Department of Defense (DOD), and the National Aeronautics and Space Administration (NASA). This program is currently installed on the Intelligent Gateway Processor of the Technology Information System (TIS/IGP) at the Lawrence Livermore National Laboratory and is under evaluation by the TIS user community from remote terminals by telephone dial-up, over TYMNET, and the ARPA computer network. Users are individually authorized for automated access to specific information centers, and use standard commands for the downloading, compilation, and online review of citations in a common format. Previously reported post-processing capabilities have been further expanded, permitting: (1) Online citation review, categorization, and addition of new data elements; (2) Disassembly and re-assembly of citations; (3) Statistical analysis of data field contents; (4) Cross-correlation of data field contents; and (5) Concordance generation. In addition, the new two-pass interpreter for the post-processing program permits: the transformation of abbreviated data field names into English names preferred by each agency, the statistical analysis of the density and completeness of data fields in selected sets of bibliographic citations, the elimination of redundant citations (using user-specified criteria), and trend analysis. The latter is a powerful tool for the exploration of time-dependent characteristics in a particular field of research, of an organization, or for an author. Graphical displays of publication rates as a function of time and the normalized statistics of terms used in the description of the work, can be used to signal new directions of ongoing research and the intensity of its support.

* Work performed under the auspices of the U. S. Department of Energy by the Lawrence Livermore National Laboratory under contract number W-7405-ENG-48.

CONTENTS

	Page
INTRODUCTION	1
1. THE "TIS" INTELLIGENT GATEWAY PROCESSOR	2
2. AUTOMATED ACCESS PROCEDURES	2
3. DOWNLOADING OF INFORMATION	4
4. POST-PROCESSING CAPABILITIES	5
Translate	5
Merge	6
Stat	6
Analyze	7
Review	7
Concord	8
Permute	8
Cross	9
Plot	9
Display	10
5. FUTURE WORK	10
ACKNOWLEDGMENT	12
REFERENCES	13

INTRODUCTION

More than 2,225 bibliographic and numeric data files are presently available from approximately 327 different online information vendors. [1] This profusion of sources makes it increasingly difficult to identify relevant bibliographic citations in a unified manner and to extract the meaningful scientific and technological (S&T) intelligence.

Most online bibliographic information is still being marketed by offline printing following a search. At best, these citations are displayed or printed in chronologically reversed order. A retrospective search carried out in this conventional manner may present the end user with stacks of printouts containing redundant citations in formats that differ according to the preference of each vendor. Usually, there are no indices or contents lists to the piles of data, requiring the end user to engage in a tedious manual review. Much of the information may be left unused. [2-4]

Even among the increasingly popular 'front-end systems' now being marketed, the user still only has a limited number of features to choose from such as downloading and file creation (i.e., SciMate, InSearch, CONIT). None of the front-end systems we have studied provide for the active manipulation of the data to permit full extraction of their information content.

At the Lawrence Livermore National Laboratory (LLNL) we have developed self-guiding programs by which some of these tasks can be carried out automatically using the dedicated Intelligent Gateway Processor (IGP) of the Technology Information System (TIS). This system, under development and in use since 1975, contains an expanding master directory to databases of other information centers. Authorized users are connected to the named information center automatically and can download desired information to TIS. The resulting files can then be post-processed by programs that permit online review, the display of statistics, the creation of indices and concordances, and text analysis. In view of the uncertain legal and monetary implications of these powerful procedures, we have limited our applications to the online information systems of the Department of Energy (DOE/RECON), Department of Defense (DOD/DROLS), and the National Aeronautics and Space Administration (NASA/RECON). In the future, we plan to explore extensions to other federal and commercial information systems as well.

1. THE "TIS" INTELLIGENT GATEWAY PROCESSOR

The TIS/IGP is a new-generation, dedicated information machine in operation since 1976. [5-9] It supports several user communities which create their own preferred view of their resources consisting of internal programmatic files, application programs, models, and different means of communications. When additional information, numeric data, or computational power are needed, TIS connects to other external information centers and computers in an automated and controlled manner. Authorized users simply specify the target name of the desired resource. In addition to these shared, internal and external resources which are organized and controlled by the system administrator, (who is responsible for the resources of a programmatic user community), each user is assigned his private work space. (See Fig. 1.)

This type of flexible and extensible work environment, and the TIS interactive tools for analysis, synthesis, and post-processing of information and data are needed to increase productivity and to transfer technology to other R&D programs and to industry in a cost-effective manner. TIS gives this capability to each user. Users also can define and create their preferred view of personal resources, e.g., data files, reports, graphs, and communications by activating self-guiding (and reasonably forgiving) menu-driven routines. Initially, the results of any work belong to the user/creator. It requires a PERMIT command to share newly generated resources with someone else, a group of co-workers, or to release them for common use through the Resource Administrator of a particular TIS user community. Since much of the daily work in R&D is being documented on electronic work processors (WP), we have also established the capability of linking with several of these WP machines for transfer of information and data to and from TIS in the TTY-equivalent mode, and in some cases in their internal WP format.

2. AUTOMATED ACCESS PROCEDURES

We are presently developing automated and transparent access procedures to different information centers as an integral capability of the Intelligent Gateway Processor (IGP). Users of TIS may consult the TIS master directory for the availability of programmatic resources stored internally for their use on TIS, or for those made available to them by automated access to external information centers.

Each external information center is qualified online on TIS. This establishes its accredition, the availability and cost of its databases, and an annotated description of salient commands and prevailing up-times. This information is extracted from a foreign host by periodic transfer to TIS. A consultation of potentially available resources by the TIS user community prior to their use saves time and communication costs. Access to federal and commercial resources is granted to TIS users on an individual basis by the Resource Administrator. Authorized users gain access to other information centers simply by giving the command CONNECT, followed by the target name of the desired resource: e.g. CONNECT DOE/RECON will establish access to DOE/RECON at 1200 baud.

Alternately, users may specify the TIS option number of the desired resource, which is part of each online menu. In either case, users need not be familar with telephone dial-up numbers, accounts, passwords, or the peculiar protocols of foreign information centers.

Main user communities of the Technology Information System establish their own views of their internal and external programmatic resources in a self-guiding manner,

Conventional user access:

TIS-intelligent gateway access:

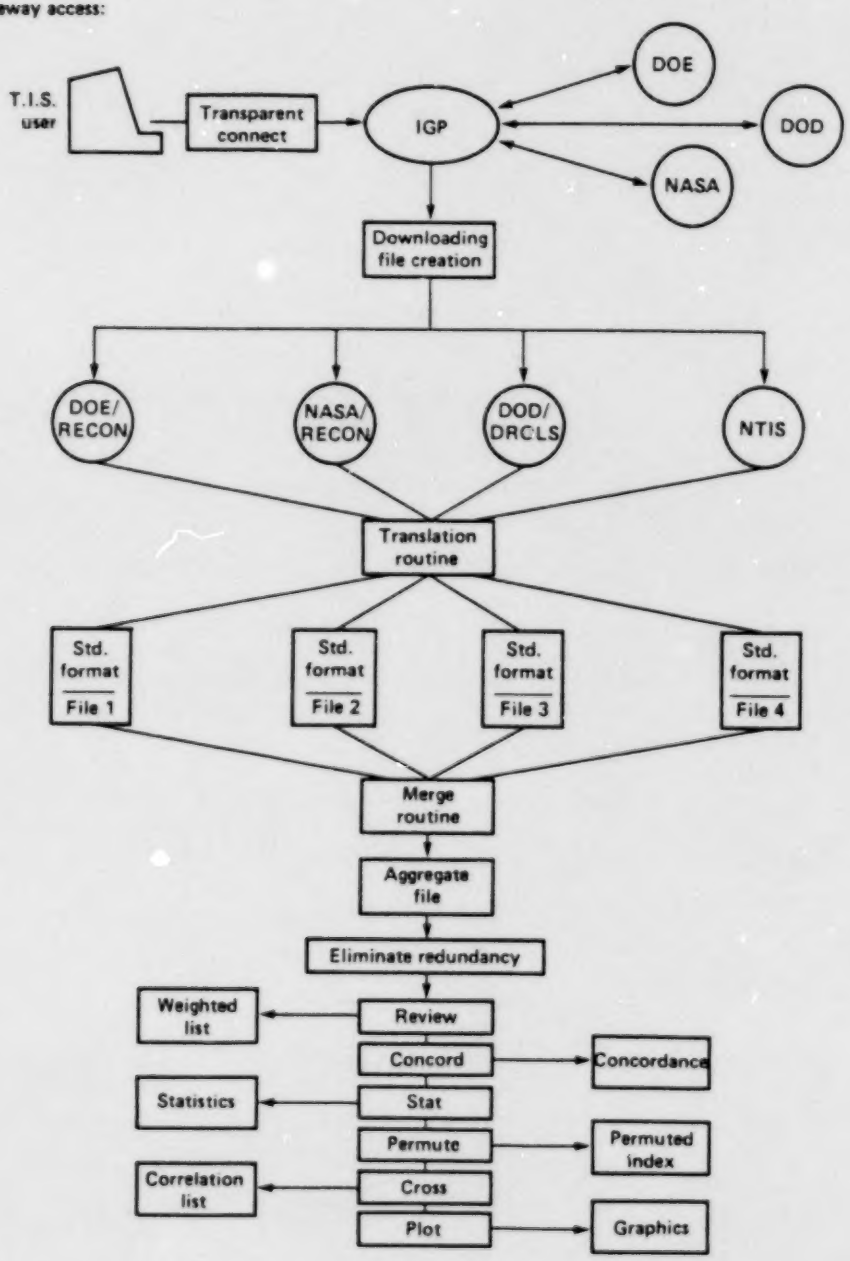


Fig. 1

BEST COPY AVAILABLE

without programmer intervention. Resource Administrators assigned to each user community control also grant access rights on a need-to-know/need-to-use basis. Individual users see only those resources (e.g., datafiles, interactive models, graphs, and reports) to which they have access. An exception today is the directory to external resources which we advertise to promote their use. When a user is finished using a TIS-provided external resource, his access rights can be removed by the TIS Resource Administrator and no change of passwords is required since none were disclosed.

The DIAL command provides an equally powerful, but user-controlled, method for accessing other information centers and computers. In this case, the user is prompted to specify the telephone number, baud rate, parity, and other communication parameters, i.e.

	Telephone	Baud rate	Parity	Duplex
DIAL	[number]	[-300, -1200, -2400]	[-o, -e]	[-h, -f]

TIS then establishes the communication similar to an automated telephone dialer. However, users have to provide their own accounts and passwords for login on the external host machine. Such procedures can be saved for personal, routine use. TIS maintains detailed audit files of all transactions and authorizations granted to users.

When an account with another information center is opened for TIS, the vendor bills TIS at LLNL, which in turn deducts the costs from the responsible programmatic accounts. When users establish their own accounts with other information centers and use them via TIS, they are billed directly by the vendors who cannot distinguish by what means the user accessed and used their information center.

3. DOWNLOADING OF INFORMATION

Users can initiate downloading of information from geographically distributed centers to TIS in two ways: First, the SAVEON option permits extraction of information when used with the CONNECT command discussed previously. In this case, all the information seen on the screen during one session is placed into a user-named file. Approximately 100 citations with abstracts can be extracted and saved in 10 minutes at 1200 baud by asynchronous telephone dial-up. Faster transmission is possible with the new 2400 baud modems and with 9600-baud, dedicated and conditioned synchronous lines. Second, the DIAL command permits extraction and downloading into one, or into several, individual, user-named files that can be opened and closed at liberty by special control characters during a session, e.g.

ESCAPE CTRL-A - prompts the user for a file name and saves the viewed information therein. An additional ESC-CTRL-A closes the file. If the file already exists, the new information is appended to permit progressive creation of a cumulative subject datafile.

ESCAPE CTRL-B - sends a local file from TIS to a remote machine. This has particular importance when downloaded and saved information is to be transferred to more powerful computers for analysis, or is to be shared with someone else via TIS.

ESCAPE CTRL-Z - permits suspension of communications from an information center, return to the TIS gateway at Livermore, and establishes automated connection to yet other, geographically distributed information centers. Users can thus be connected to more than one informaton center at the same time and compare their resources on the TIS gateway.

Similar commands permit the user to stop the viewing, and/or saving of information and to analyse the data on the gateway or on some other computer, for extraction of the best available data and cost-effective decision making. [10]

4. POST-PROCESSING CAPABILITIES

When a retrospective search is carried out for a new field of interdisciplinary research, it is not unusual to obtain thousands of citations from different information vendors, in different formats, with redundant overlap. [11] The IGP offers a list of programs, or menu-driven routines, for the manipulation and re-organization of downloaded files, enabling the user to gain added insight from the data at hand. These include:

TRANSLATE	-	translates citations to a standard format.
MERGE	-	compiles translated files from different sources into one file.
STAT	-	creates a statistical profile of citations.
ANALYZE	-	provides bibliographical text analysis.
REVIEW	-	permits online evaluation of citations for relevancy.
CONCORD	-	creates indexes by author, subject, descriptors, etc.
PERMUTE	-	issues multi-term statistics of the text in selected data fields.
CROSS	-	cross-correlates the contents of data fields.
PLOT	-	shows the number of citations by year in a graph.
DISPLAY	-	presents the contents of any file on the CRT screen.

The TRANSLATE routine transforms downloaded bibliographic citations from different sources into a standard format is a basic capability of the TIS/IGP. It removes the peculiarities from citations characteristic of a particular bibliographic origin by establishing a common, consistent format representation. This allows the user to concentrate on the information contents of each citation without concern about their disparate origin.

Specifications for the translation of citations from different formats can be defined and altered by the end-user in an interactive manner. Once submitted, the specifications become part of the user-specific translation tables for a particular bibliographic source or for a group of similar bibliographic sources. A new translation table can thus be created with little or no additional assistance. To this end, a group of English-like verbs have been created which define the actions to be taken at key points in the translation process. New verbs denoting new actions can be created as needed by the programming

staff. This allows modifications and extensions to the translation table to be made by the information specialist or end-user during his session, making him independent of programmer intervention and of changes introduced by an information center. Each part of the translator specification language describes a particular aspect of the translator. For example the term:

Source	-	Defines the start of a citation from a new bibliographc source in a compiled list.
Citation	-	Defines the start of a citation in a list.
End source	-	Defines the end of a bibliographic source.
Spurious	-	Defines remedial actions for inadvertent damage to, and page breaks in, downloaded files.

The translator is table driven, allowing new bibliographic sources to be added without impacting the rest of the translator system. Conditions and data elements that cannot be identified are flagged as unrecognized without loss of information. In summary, the IGP translator:

- provides the common format for all other post-processing routines.
- can be defined interactively by its specification language.
- automatically identifies a bibliograpic source.
- generates a consistent, standard format for subsequent work.
- provides an audit trail for the origin of citations encountered.
- manages, issues warnings, and recovers from errors in communications and format.

The MERGE routine of the IGP translator allows the aggregation of several files in the common IGP format to be merged into a composite file for subsequent post-processing. Files may be located in the user's account, or may be in other accounts of the IGP where read/write privileges have been granted.

The statistical analysis routine STAT establishes for the user the density and completeness of data elements of the file containing the relevant citations. These statistics are shown online to permit the user to select those data elements that exist in a majority of citations (e.g., accession number, title, author, and descriptors are usually present and are shown as absolute counts and as percentages; here 100%). This aids the systematic analysis of information by establishing a complete set for subsequent, internally consistent post-processing. Initial statistical results are displayed as shown:

```
STATISTICS FOR FILE: doe on Fri Jul  6 15:17:50 1984

 No.  Count  Per Field-Name            Description

  1:*    80 100% <ABSTRACT>          = Recon: <ABS     >
  2:*    80 100% <ACCESSION NO.>
  3:*    80 100% <ANN J>             = Recon: <SJA     >
  4:*    80 100% <BIS>               = Recon: <BIB     >
  5:*    80 100% <CATEGORIES>        = Recon: <SCC     >
  6:*    80 100% <CO OF AUTH>        = Recon: <LWC     >
  7:     80 100% <CO OF PUBL>
  8:*    80 100% <DATE>              = Recon: <PDD     >
  9:*    80 100% <DESCRIPTORS>       = Recon: <GEN     >
 10:     80 100% <DOCUMENT NO>       = Recon: <DSN     >
 11:     80 100% <ISSUE>             = Recon: <SJI     >
 12:     80 100% <PRIMARY CAT>       = Recon: <PCC     >
 13:     80 100% <TYPE>              = Recon: <PTC     >
 14:     80 100% <UPPOSTED DESC>     = Recon: <DSC     >
-------------------------------------------------------------------
 15:     54  67% <PAGE NO>           = Recon: <PGM     >
 16:     54  67% <TITLE(MONO)>       = Recon: <TLM     >
```

BEST COPY AVAILABLE

This listing may be modified by the user for a display of a selected subset of citation fields, either by frequency or in alphabetical order.

The ANALYZE routine provides the user with a means to structure a quick bibliographical text analysis. The default mode reformats each citation of the selected file into a simple format containing only the title, author, accession number, and a partial display of the abstract. Screen placement and indentation eliminate the need for field labels and give the user a simple, uncluttered view of each citation. In addition, the user may modify the formatting to include his own field parameters - i.e., descriptors, input data, etc.

```
Nuclear-safety criteria and specifications for space                84R00001752
  nuclear reactors

    1982   USDOE Assistant Secretary for Nuclear Energy,
           Washington, DC. Office of Space Nuclear Projects

           USDOE Assistant Secretary for Nuclear Energy,
           Washington, DC. Office of Space Nuclear
           Projects.

    The policy of the United States for all US nuclear
    power sources in space is to ensure that the
    probability of release of radioactive material and the
    amounts released are such that an undue risk is not
    presented, considering the benefits of the mission. The
    objective of this document is to provide safety
    criteria which a mission/reactor designer can use to
    help ensure that the design is acceptable from a
    radiological safety standpoint. These criteria
    encompass mission design, reactor design, and
    radiological impact limitation requirements for safety.
```

The REVIEW command permits online determination of relevancy. Citations are shown on the screen reformatted by accentuation and indentation of titles, authors, and abstracts. This renders them more readable than citations commonly offered by information centers. The viewer may keep or discard any citation shown and assign his own category and relevancy code. He may add comments, order the full-length text, and define and fill new data fields for numeric and/or administrative purposes. Retained and annotated citations are saved in new user-named files. Fields defined during the review process can subsequently be used with other fields for post-processing.

```
Gaseous fuel reactors for power systems                             84B00001756

    129630208

           Helmick, H.H.; Schwenk, F.C.

           Society of Automotive Engineers, Inc.,
           Warrendale, PA

    The Los Alamos Scientific Laboratory is participating
    in a NASA-sponsored program to demonstrate the
    feasibility of a gaseous uranium fueled reactor.  The
    work is aimed at acquiring experimental and theoretical
    information for the design of a prototype plasma core

Rest of abstract? (y or n): n

Keep citation? (y or n): y
Local options? (y or n): y

Category:  systems
Relevancy: 9
Comment:   Relates to systems only
Order text (y/n): n
Saved.
```

The CONCORD routine generates concordances by author, descriptors, corporate author, or country of origin and yields succinct listings of bibliographic citations in a particular field. These alphabetical indices are similar to those commonly produced as look-up tables for authors or subjects. In this case, they are created at the pleasure of the user, online, on the contents of any citation field.

Ablekov, V.K.			
1982	Avduyevskiy, V.S.; Grishin, S.D.; Leskov, L.V.; Ablekov, V.K.; Yevich, A.F.	Prospects for power engineering in space	83R0110459
Alemany, A.			
1983	Thibault, J.P.; Jousselin, F.; Alemany, A.; Dupas, A.	Metal gas MHD converter development plans	83C0149919
Anderson, R.V.			
1983	Anderson, R.V.; Bost, D.; Determan, W.R.	Space-reactor electric systems: subsystem technology assessment	83R0074813
1983	Anderson, R.V.; Bost, D.; Determan, W.R.; Harty, R.B.; Katz, B.;	Space reactor electric systems: system integration studies, Phase 1 report	83R0084623

The PERMUTE routine provides a highly significant aspect of post-processing by the time-dependent change, or momentum, of words used to describe a particular set of relevant citations. It can be derived in part from the statistics of its permuted descriptive terms. The PERMUTE command of our post-processing routines provides this option by counting the number of times a specified term appears in the message-carrying fields of citations, e.g., their titles, abstracts, descriptors, categories, etc. This is done by analyzing single and compound expressions containing up to four terms (i.e., solar energy conversion experiments). All compound expressions of this type appearing in the selected data fields are presented to the viewer online and are ordered alphabetically or according to their frequency of occurence. The tables below show this for two fields - descriptor and author.

```
6 •HEAT PIPES--performance testing
6 •HEAT PIPES--specifications
6 FEASIBILITY STUDIES
6 HEAT PIPES
6 SPACE FLIGHT
5 •SPACE POWER REACTORS--reactor materials
5 •SPACE POWER REACTORS--shielding
5 •SPACE POWER REACTORS--technology assessment
5 •SPACE PROPULSION REACTORS--heat pipes
5 ELECTRIC POWER
5 SPACE VEHICLES
4 •REACTOR CORES--specifications
4 •SPACE POWER REACTORS--feasibility studies
4 •SPACE POWER REACTORS--planning
4 PERFORMANCE
4 RADIATORS
4 RELIABILITY
4 SPACECRAFT POWER SUPPLIES
4 VERY HIGH TEMPERATURE
```

BEST COPY AVAILABLE

```
10 Buden, D.
 7 Ranken, W.A.
 4 Bennett, G.L.
 4 Keddy, E.S.
 3 Determan, W.R.
 3 Ernst, D.M.
 3 Girrens, S.P.
 3 Koenig, D.R.
 3 Lundberg, L.B.
 3 Martinez, H.E.
 3 Merrigan, M.A.
 2 Anderson, R.V.
 2 Angelo, J.A. Jr.
 2 Bost, D.
 2 Botts, T.E.
 2 Harty, R.B.
 2 Layton, J.P.
 2 Lillie, A.F.
 2 Lombardo, J.J.
 2 Palmer, R.G.
 2 Powell, J.
```

The CROSS routine generates cross-correlations of expressions contained in any two fields of citations and can provide new insight to their content. For example, by cross-correlating authors, we can see at a glance who has been working with whom. A cross-correlation of the author field with the descriptor field shows, in alphabetic order, the statistics of indexing terms assigned to the work of a particular person for his entire professional career. When carried out in yearly increments, the CROSS routine can be used to judge the change of emphasis in the published works of a scientist, or an organization, as a function of time.

```
Barlow, T.W.      11* Barlow, T.W.
                   1  Burrows, C.R.
                   2  Chiao, T.T.
                   2  Cornell, E.P.
                   3  Crothers, W.T.
                   2  Frank, D.N.
                   3  Kulkarni, S.V.
                   1  Reimers, E.
                   2  Rinde, J.A.
                   2  Turnbull, F.G.
                  31  TOTAL

Bauer, W.H.        2* Bauer, W.H.
                   1  Brobeck, W.M.
                   1  Younger, F.C.
                   4  TOTAL

Beachley, N.H.    10* Beachley, N.H.
                   1  Dietrich, A.
                   9  Frank, A.A.
                   1  Hartar, R.
                   1  Jamzadeh, F.
                   1  Lau, K.
                   1  Otis, D.R.
                   1  Stockman, D.
                   1  Volz, T.
                  26  TOTAL
```

Author-author correlation.

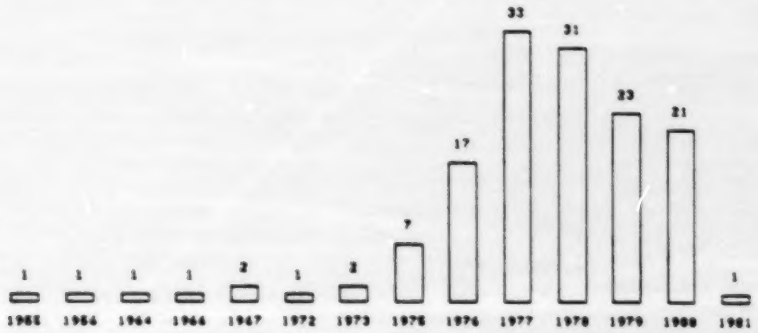
```
Davis, B.          1  COMMERCIALIZATION
                   2  FLYWHEEL ENERGY STORAGE
                   3  FLYWHEEL-POWERED VEHICLES
                   2  FLYWHEELS
                   3  HYBRID ELECTRIC-POWERED VEHICLES
                   1  OPERATION
                   2  PERFORMANCE TESTING
                   3  PERFORMANCE
                  25  TOTAL Descriptors

General Electric Co.,   1  COMPUTER CALCULATIONS
  Schenectady, NY(USA)  1  COMPUTERIZED SIMULATIONS
  Corporate Research and 2  CONTROL EQUIPMENT
  Development Dept.     1  COST BENEFIT ANALYSIS
                        2  DESIGN
                        1  ELECTRIC GENERATORS
                        1  ENERGY STORAGE SYSTEMS
                        2  EXPERIMENTAL DATA
                        2  FABRICATIONS
                        3  FEASIBILITY STUDIES
                        3  FLYWHEEL ENERGY STORAGE
                        4  FLYWHEEL POWERED VEHICLES
                        1  FLYWHEELS
                        3  HYBRID ELECTRIC-POWERED VEHICLES
                        1  LIFE CYCLE COST
                        2  PERFORMANCE TESTING
                        2  RESEARCH PROGRAMS
                        1  STEELS
                        1  WELDING
                       34  TOTAL Descriptors
```

Author-descriptor correlation.

A PLOT of the publication rate for a particular search topic, an institute, or author provides an immediate indication of the effort, i.e., growth, in their field of activity. In most cases, as shown in the example below, we see an apparent decline after a sharp rise in the publication rate. This decline is predominantly the lag between the publication of the primary literature and its entry into the secondary online database holdings. To appreciate an above-average increase in publications in a particular field, one has to compare a specific rate increase with the total annual increase of publications. In the sciences, this rate is now about 13%.

BEST COPY AVAILABLE

DOE/RECON Data Analysis Program
Number of Publications per Year



This DISPLAY option allows users to display, view, and verify the contents of any file during the course of their post-processing work. It is most useful prior to merging the file and as an ad hoc tool for inspection of post-processed lists.

5. FUTURE WORK

The statistical analysis of the contents of a data element as a function of time is a new and powerful option we plan to develop for our user community in the near future. In the example shown below (developed with the aid of available post-processing routines and limited to their analysis by yearly increments), the trend analysis was conducted on citations retrieved from the DOE/RECON, NASA/RECON, and DOD/DROLS databases from a search for citations dealing with "computers" and "security." Note the growth of publications over the years and the appearance of new terms that characterize a rapidly expanding technology.

When fully implemented, trend analysis routines will be automated and self-guiding. We plan to provide absolute counts of terms and the percent change derived from statistics of expression used in the descriptive fields of citations. This should permit us to speculate and/or predict the direction of research and technological development as a function of time.

TREND ANALYSIS - A TOOL FOR THE PREDICTION OF FUTURE R&D

	Number of Citations:	1	1	3	1	2	5	4	9
	DESCRIPTORS	1975	1976	1977	1978	1979	1980	1981	1982
1975	COMPUTER NETWORKS	1	1	3	1	2	5	4	9
	OPERATION	1					1		
	PDP COMPUTERS	1							
	PROGRAMMING	1					1		2
	RELIABILITY	1	1				1		1
	SECURITY	1	1	3	1	2	5	4	9
1976	COMPUTERS		1	1			2	1	3
	CRIMINOLOGY		1						
	DESIGN		1		1		1		
	EDUCATION		1						
	ENERGY CONSERVATION		1						
	ENERGY POLICY		1						
	MANAGEMENT		1	2			1	1	
	MATHEMATICAL MODELS		1						
	MEETINGS		1				1		1
	MEMORY DEVICES		1	1			1		
	MICROPROCESSORS		1						
	MILITARY EQUIPMENT		1						
	PUBLIC HEALTH		1						
	SEMICONDUCTOR DEVICES		1						
	TECHNOLOGY ASSESSMENT		1						
	TECHNOLOGY UTILIZATION		1						
1977	COMPUTER OUTPUT DEVICES			1					
	DATA TRANSMISSION			1				3	
	FORECASTING			1					
	OPERATING COST			1					
	SPACIAL DISTRIBUTION			1					
	US ERDA			2					
1978	KERNELS				1				
1979	ALARM SYSTEMS								
	ALGORITHMS					1			1
	AUTOMATION					1			
	COMPARATIVE EVALUATIONS					1			
	COST					1			1
	DATABASE MANAGEMENT					1			1
	MONITORING					2			
	NUCLEAR MATERIALS MGMT					1			
	NUCLEAR WEAPONS					1			
	PERFORMANCE					1			1
	RADIOACTIVE MATERIALS					1			
	STORAGE					1			
1980	AMDAHL COMPUTERS						1		
	COAXIAL CABLES						1		
	COMMUNICATIONS						1		
	COMPUTER GRAPHICS						1		
	DATA TRANSMISSION						2		
	ENFORCEMENT						1		
	INSPECTION						1		
	JOINING						1		
	LEGAL ASPECTS						1		
	OPERATION						1		
	PERSONNEL						2		1
	REAL TIME SYSTEMS						1		
	SANDIA LABORATORIES						1		1
	TRAINING						1		
1981	DATA PROCESSING							1	
	MANUALS							1	
	SECRECY PROTECTION							1	
	SYNCHRONIZATION							1	
1982	COMPUTER CODES								1
	COMPUTERIZED SIMULATION								1
	EQUATIONS								1
	OPTIMIZATION								1
	PRODUCTIVITY								1
	RECOMMENDATIONS								1
	RISK ASSESSMENT								2
	SABOTAGE								1
	SAVANNAH RIVER PLANT								1
	SECRECY PROTECTION								3
	SIMULATION								1
	US DOE								1

ACKNOWLEDGMENT

The post-processing routines for bibliographic citations of the DOE/RECON, NASA/RECON, and DOD/DROLS information systems were developed under joint sponsorship of the online services of their respective agencies. Interactive routines available now permit the aggregation of bibliographic information from these three leading sources of federally sponsored research. These new computer tools yield new, fascinating insight for ongoing programs, and can be used to predict future growth and emphasis in emerging technologies.

This development - post-processors and translator - is part of a larger activity among the agencies to integrate their immense S&T resources of bibliographic information and numeric data in an Information Center Network (ICN). This network is planned to consist of a number of Intelligent Gateway Processors, rather than over hard-wired interconnections alone. These gateways would link those in need of up-to-date S&T information with the large federal information centers, and with each other in a unified manner, leaving the heterogeneous resources at their present geographically distributed locations. The TIS/IGP program supports the congressional mandate for Technology Transfer under PL-96-480 and DOE Order 5800.1.

The work has been carried out by the staff of the Technology Information System (TIS) project at the Lawrence Livermore National Laboratory (LLNL), and by system analysts from Control Data Corporation and several consulting firms under contract to LLNL/TIS.

REFERENCES

1. R. N. Cuadra, D. M. Abels, and J. Wagner, Directory of Online Databases (Cuadra Associates, Inc., Santa Monica, CA, 1984), Vol. 5, No. 3, Spring 1984.
2. J. K. Burnam, Users' Needs for Repacking and Reuse of Information, Information Services and Use (North Holland Publishing Co., 1982), Vol. 1, pp. 359-366.
3. W. A. Bollinger, V. E. Hampel, T. P. Murphy, User Requirements for Post-Processing of Bibliographic Information, Lawrence Livermore National Laboratory, Livermore, CA, UCRL-10014 (February 1984).
4. V. E. Hampel, Fact Retrieval for the 1980's, Proceedings of the Technical Information Panel (TIP), NATO/AGARD Conference (Munich, West Germany, AGARD-CPP-304, 1981); also published by the Lawrence Livermore National Laboratory, Livermore, CA, as UCRL-85749 (July 1981).
5. V. E. Hampel, Intelligent Gateway Computers for Nationwide Information Systems of the Future, Post conference proceedings of the World Future Conference (Washington, D. C. July 19-21, 1982).
6. V. E. Hampel, W. A. Bollinger, C. A. Gaynor, B. E. Hegemann, R. D. Sanner, and Y. Wolman, International Networks for Material Properties, Presented to the Ninth International CODATA Conference (Jerusalem, Israel, June 24-28, 1984); also published by the Lawrence Livermore National Laboratory, Livermore, CA, as UCRL-90942 (June 1984).
7. V. E. Hampel, J. B. Cain, R. A. Kawin, G. Pavel, N. A. Lann, G. T. Richards, and W. S. Scott, TIS - A Focal Point for Technology Transfer, Lawrence Livermore National Laboratory, Livermore, CA, URCL-53342 (September 1982).
8. V. E. Hampel, L. E. Gallo, R. A. Kawin, V. Kopytoff, S. K. McGrogan, L. G. O'Connell, G. Pavel, J. A. Schriebman, and J. E. Swanson, Editors, Proceedings of the First DOE/LLL Workshop on the Technical Management Information System (TMIS), CONF-791258 (1979).
9. V. E. Hampel, R. A. Kawin, N. A. Lann, and W. S. Scott, TIS - An Interactive System for Technology Transfer, Proceedings of the 7th International Symposium of the Technology Transfer Society (Pasadena, CA, June 13-14, 1982); also published by the Lawrence Livermore National Laboratory, Livermore, CA, as UCRL-87703 (June 1982).
10. I. Harrison, Jr., V. E. Hampel, and R. A. Kawin, Downloading and Post-Processing of Bibliographic Information With the TIS Intelligent Gateway Computer, Proceedings of the Online '82 Conference (Atlanta, GA, November 1-2, 1982); also published by the Lawrence Livermore National Laboratory, Livermore, CA, as UCRL-88119 (September 1982).
11. R. K. Hunt, H. L. Fisher, V. E. Hampel, R. A. Kawin, and N. A. Lann, The "TIS" Intelligent Gateway Computer, An Alternative to the "Doomsday Scenario," Presented to the 4th National ONLINE Meeting at the Sheraton Center (New York, NY, April 12-14, 1982); also published by the Lawrence Livermore National Laboratory, Livermore, CA, as UCRL-88784 (February 1983).

7277c